Prediction of Traffic Complexity and Controller Workload in Mixed Equipage NextGen Environments

Paul U. Lee
San Jose State University /
NASA Ames Research Center

Thomas Prevot
NASA Ames Research Center

Controller workload is a key factor in limiting en route air traffic capacity. Past efforts to quantify and predict workload have resulted in identifying objective metrics that correlate well with subjective workload ratings during current air traffic control operations. Although these metrics provide a reasonable statistical fit to existing data, they do not provide a good mechanism for estimating controller workload for future air traffic concepts and environments that make different assumptions about automation, enabling technologies, and controller tasks. One such future environment is characterized by en route airspace with a mixture of aircraft equipped with and without Data Communications (Data Comm). In this environment, aircraft with Data Comm will impact controller workload less than aircraft requiring voice communication, altering the close correlation between aircraft count and controller workload that exists in current air traffic operations. This paper outlines a new trajectory-based complexity (TBX) calculation that was presented to controllers during a human-in-the-loop simulation. The results showed that TBX accurately estimated the workload in a mixed Data Comm equipage environment and the resulting complexity values were understood and readily interpreted by the controllers. The complexity was represented as a "modified aircraft count" that weighted different complexity factors and summed them in such a way that the controllers could effectively treat them as aircraft count. The factors were also relatively easy to tune without an extensive data set. The results showed that the TBX approach is well suited for presenting traffic complexity in future air traffic environments.
.

INTRODUCTION

The Next Generation Air Transportation System (NextGen) will overhaul the U.S. National Airspace System (NAS) to make air travel more efficient and dependable while maintaining the high level of safety that we enjoy today. NextGen is envisioned to transform the air traffic environment by introducing new enabling technologies and automation (JPDO, 2007).

In the NextGen timeframe, a key limiting factor for a significant increase in en route air traffic capacity is expected to be controller workload. Therefore, accurate estimation of controller workload has been an active field of research (e.g., Chatterji, G. B. and Sridhar, B., 2001, Kopardekar and Magyarits, 2003; Masalonis, Callaham and Wanke, 2003; Hilburn and Flynn, 2005; Histon and Hansman, 2008). Many of these efforts have focused on developing new complexity metrics to predict controller workload in NextGen. Unlike current air traffic environment, NextGen operations are expected to accommodate aircraft with much more diverse avionics and ground equipage, which makes the current method of using simple aircraft count to predict controller workload unsuitable for NextGen.

A main effort in traffic complexity research has been to develop Dynamic Density (DD) metrics (Kopardekar and Magyarits, 2003). DD metrics are a large set of complexity factors that have been identified and correlated with controller workload across a diverse set of sectors. Although combined DD metrics estimated controller workload better than aircraft count alone, many of the DD factors were not suited to predict traffic situations in real-time operational settings (Bloem, Brinton, Hinkey, Leiden, and Sheth, 2009). Some of the DD metrics exhibited large uncertainty when they were projected over a two-hour time horizon, resulting in significant fluctuations in the complexity values. In addition, DD and other complexity metrics that have been proposed over the years have a shortcoming in that they do not allow the human operators to easily interpret the data and translate them into actions that could support air traffic management initiatives.

As an alternative approach to traffic complexity calculations, Hansman and his colleagues proposed an integrated metric called Modified Aircraft Count, which identifies a set of complexity factors that are easily understood by air traffic personnel (Histon and Hansman, 2008). The Modified Aircraft Count metric assigns different weightings to different aircraft states (e.g., aircraft in conflict, aircraft with delays, etc.) and combines them into a single factor that is calibrated to represent the effective number of aircraft in a sector. The Modified Aircraft Count can be used by air traffic personnel in a way similar to how they handle today's Monitor Alert Parameter (MAP), which is based on predicted aircraft count.

Using a similar approach, a new Trajectory-Based Complexity (TBX) was developed at the Airspace Operations Laboratory at NASA Ames Research Center. TBX calculates complexity values that are readily interpretable by human operators in unique NextGen operations. The following sections, give an overview of TBX calculations and how they were used in an evaluation of one NextGen concept.

TBX OVERVIEW

Based on the lessons learned from past research, a set of feature characteristics were identified to drive our TBX calculations: 1) metrics that were comprehensible to human operators; 2) metrics whose weightings would be easy to modify for new equipage and operational environments; 3) metrics that were stable over the prediction time horizon; 4) use of trajectory-based operations to create more stable trajectory predictions over a one- to two-hour time horizon; and 5) the combination of different complexity factors into a single modified aircraft count value that could be used like MAP.

Details of the TBX calculations is extensive and beyond the scope of this paper. A full description can be found in Prevot and Lee (2011). This paper provides a basic overview of TBX and the complexity factors that are a part of the calculations.

Calculating the TBX value is based upon establishing nominal conditions in which aircraft count is a good predictor of controller workload. Then the actual conditions are compared to those nominal conditions, and the aircraft count is adjusted accordingly. Thus, TBX mimics aircraft count and the same mechanism of setting capacity thresholds using a MAP that reflects the maximum aircraft count can be used with TBX. The TBX value will be equal to the predicted aircraft count for a "nominal" sector size, a nominal number of transitioning aircraft and a nominal equipage mix, but any predicted difference to nominal operations, such as convective weather cells, different aircraft equipage mixes, unusually high levels of transitioning aircraft or traffic conflicts, causes a modification to the aircraft count, which will be reflected in the respective TBX value.

TBX calculations consist of a summation of two parts. First the *predicted aircraft count* is multiplied by adjustments for *sector size* and *predicted weather penetrations*. A "nominal sector" size is initially defined for a given airspace. An adjustment formula is defined to compute a sector size factor that will to increase the complexity value for small sectors and reduce the value for larger sectors. For weather penetrations, the predicted weather trajectory is compared against the predicted aircraft trajectory to determine whether these two intersect within a sector. A formula relating the number of predicted weather penetrations to the sectors MAP value is used to adjust the predicted aircraft count for weather.

The second part of the TBX calculations uses a weighted sum of other complexity factors that further modify the aircraft count. These factors include the *number of transitioning aircraft*, the *equipage mix*, and the *number of conflicts* that occur in a sector.

The nominal values for these factors are fairly easy to estimate in TBX calculations. For a given sector, one can collect past traffic and weather data to see the peak aircraft count, the number of transitioning aircraft, etc. that the sector nominally handles in different situations. For NextGen related factors, such as the Data Comm equipage mix, a good database does not exist to estimate the values. However, it was not too difficult to converge on a reasonable set of parameters even for factors with no historical data. The nominal values can first be generated by running a sample traffic scenario in the target airspace in fast-time simulations or eliciting inputs from subject matter experts to estimate the values. Targeted simulation runs with actual controllers can be conducted to validate the parameter estimations and adjust them further if necessary.

The two parts are combined such that all complexity factors result in a single TBX value, a modified aircraft count that can be used to estimate controller workload and alert operators when it exceeds a MAP value like today. Figure 1 shows a prototype display of a traffic load table that shows both the predicted aircraft count and the corresponding complexity values in fifteen-minute increments.

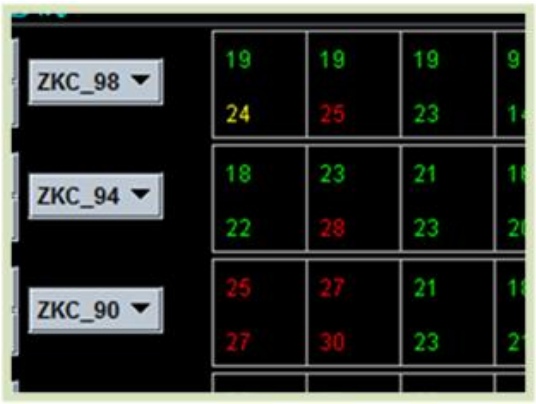

Figure 1. Traffic Load Table with predicted aircraft count (on top) and TBX value (on bottom) for sectors 98, 94, and 90. Red numbers indicate that the value exceeds the yellow maximum aircraft or complexity threshold (set at 24). Green numbers are below the set threshold.

A key advantage of using TBX values in place of MAP is that TBX can incorporate the impact of different aircraft equipage, weather, and other complexity factors have on controller workload. Unlike most other complexity metrics, however, controllers and supervisors can use TBX in a similar way as they use MAP values today for setting the sector traffic thresholds. The following sections describe how TBX values were used in a Multi-Sector Planning (MSP) mixed equipage study.

MULTI-SECTOR PLANNING IN MIXED EQUIPAGE AIRSPACE

The MSP concept addresses demand/capacity imbalances by allowing non-controller positions (e.g., traffic management coordinators or area supervisors) to look ahead 30-90 minutes and create trajectory-based solutions to weather or congestion problems across multiple sectors.

Prior studies have been conducted to evaluate the potential benefits of the MSP concept in a full Data Comm environment. The results from those studies suggested that it was unclear whether Data Comm was required for MSP. Therefore, a follow-up study was conducted in 2010 to test the feasibility and benefits of MSP in a mixed equipage environment (Smith, et al., 2011).

Detailed results of the study are in Smith, et al. (2011). In this paper, described results will only be related to the TBX values. To that goal, the test airspace, experimental conditions, TBX calculations, and the results related to TBX calculations will be briefly described below.

Airspace and Participants

The simulation airspace included four high altitude test sectors (i.e., sectors 94, 98, 29, and 90) in Kansas City Center (ZKC) and the surrounding airspace (see Figure 2). Test sectors were divided into two areas: ZKC-North (sectors 94 and 98) and ZKC-South (sectors 29 and 90). Each test sector was staffed with a team of radar and radar associate controllers. The floor of the simulation airspace was FL290.

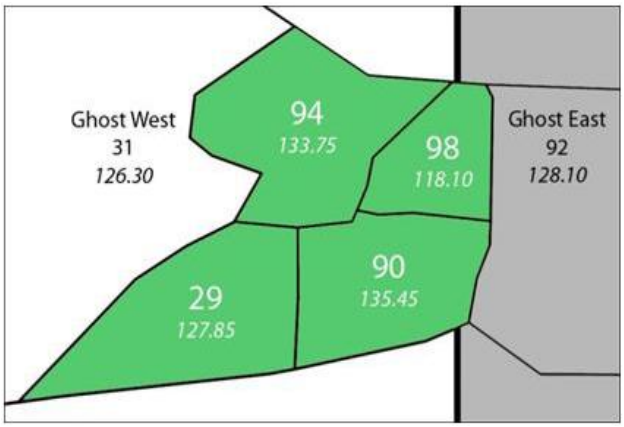

Figure 2. Test sectors 94, 98, 29, and 90 in Kansas City Center

The main focus of the MSP study was on the supervisory and traffic management positions, collectively comprising the multi-sector planning team. The ZKC traffic management team coordinated with the surrounding facilities and managed inbound and outbound flows to address problems within ZKC and the adjacent facilities. Four FAA supervisory personnel were recruited to conduct MSP operations, two area supervisors and two Traffic Management Coordinators (TMCs).

The MSP team also managed the traffic and controller workload through the test sectors that are the focus of our TBX analysis. Four radar and four radar associate positions were staffed by two additional FAA supervisory personnel who were certified on the radar position and six retired controllers from Oakland Center (ZOA). The FAA participants staffed the radar positions for sectors 94 and 29 while the ZOA retired controllers staffed the remaining radar and radar associate positions. All of the simulated aircraft were flown by simulation-pilots, who are active commercial pilots or students from the Aviation Department at San Jose State University.

TBX Assumptions for the Test Airspace

In this study, TBX values were calculated in the following manner. First a maximum aircraft threshold value of 24 was picked when 100% of aircraft were Data Comm equipped. The threshold value of 24 was selected based on past human-in-the-loop simulation studies involving 100% Data Comm equipped aircraft. The total number of aircraft that a controller could handle with Data Comm was greater than without because each Data Comm equipped aircraft reduced controller workload associated with reroutes and radio frequency communications.

Since the current day MAP or the maximum number of aircraft in the same sectors without Data Comm averaged around 18, the workload associated with controlling an unequipped aircraft was expected to be about 33% greater than that of controlling a Data Comm equipped aircraft. Independent feedback from subject matter experts supported the estimated increase of workload by 33% to be a reasonable assumption.

It was also assumed that the nominal conditions would not have aircraft penetrating weather and that a nominal sector size was 100 x 100-square nautical miles. Based on past study results, a nominal number of transitioning aircraft and a nominal number of conflicts for each test sector were set. All aircraft were also equipped with Automatic Dependent Surveillance-Broadcast (ADS-B) and a Flight Management System (FMS) to support the assumed NexGen technologies.

Experiment Design

The experiment used a within-subjects design. Each participant was designated as a radar controller or radar associate of a given sector. They managed various traffic scenarios under three mixed Data Comm equipage conditions: *10%, 50%,* and *90%* of aircraft equipped with Data Comm.

The experimental test scenarios combined two convective weather patterns (W1 and W2) across different traffic patterns. Weather patterns W1 and W2 included convective weather within the ZKC test sectors and in the surrounding airspace (Chicago or Indianapolis Centers). The weather patterns were superimposed on different traffic patterns with varying sector complexities.

Each of the traffic and weather scenarios was further modified to include different ratios of equipped (Data Comm) to unequipped aircraft in the test sectors under different experimental conditions. A total of eight scenarios with unique traffic, weather, and Data Comm equipage combinations were used for the study.

In this study, the multi-sector planning team rerouted the aircraft out of the test sectors whenever the TBX values / modified aircraft count were predicted to be above the maximum threshold defined for the study. Since unequipped aircraft contributed more to the TBX values than equipped aircraft, 10% Data Comm equipped scenarios often resulted in greater TBX values but fewer aircraft than 90% Data Comm equipped scenarios. The following section will show the resulting controller workload when the overall aircraft count and TBX values diverged as described above.

RESULTS

The main hypothesis in this paper is that in NextGen operations, predicted aircraft count may no longer correlate well with controller workload and TBX calculations could better predict workload by weighing a set of complexity factors that represent the impact of those factors on controller

workload more accurately. To test this hypothesis, analysis of the relationship between controller *workload*, *aircraft count*, and the *TBX* value were conducted.

Subjective controller workload ratings were obtained through prompts displayed on the controller display at five minute intervals throughout the simulation runs. Participants rated their perceived workload based on a 6-point rating scale, with 1 to 2 ratings as "low workload", 3 to 4 ratings as "medium workload", and 5 to 6 ratings as "high workload".

Figure 3 shows the subjective workload ratings of the radar controllers, averaged across all participants and runs. The average workload was the highest for the 10% Data Comm equipage condition and the lowest for the 90% Data Comm equipage condition. The results suggest that the aircraft equipped with Data Comm require less work from the controllers compared to the unequipped aircraft.

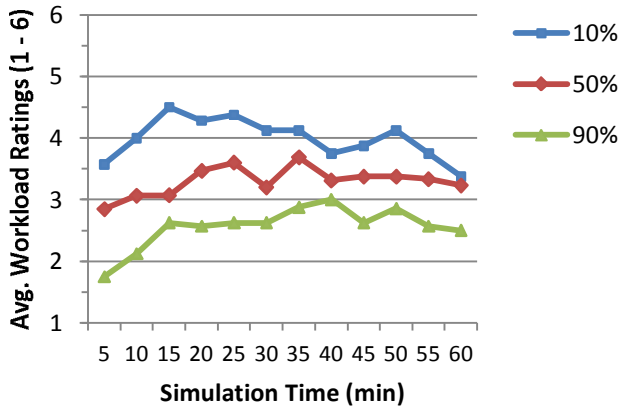

Figure 3. Average workload ratings over time for the different Data Comm mixed conditions across all test sectors and runs

Figure 4 summarizes the average aircraft count, TBX values, and workload ratings across all test sectors and simulation runs. Consistent with Figure 3, Figure 4 shows that the average workload ratings were the highest for the 10% Data Comm equipage condition (M = 4.0, SD = 0.6), lowest for the 90% condition (M = 2.6, SD = 0.8), and the 50% condition somewhere in between (M = 3.4, SD = 0.8).

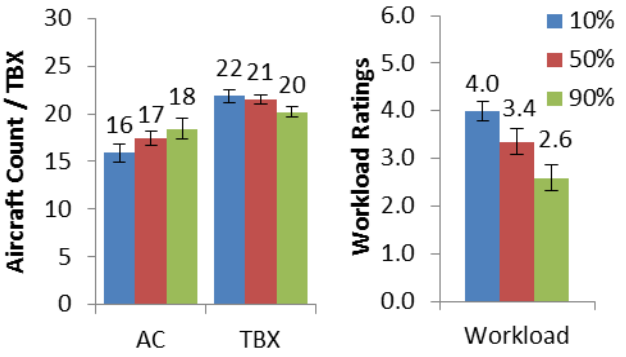

Figure 4. Average aircraft count (AC), TBX values, and workload ratings for the different Data Comm mixed conditions across all test sectors and runs

Using pure aircraft count as a predictor of workload, it would be expected that the 10% condition would have higher aircraft counts than the 50% and 90% conditions. However, the opposite was true. The average aircraft count was the lowest in the 10% condition (M = 16, SD = 2.6), highest in the 90% condition (M = 18, SD = 3.0), and in the middle for the 50% condition (M = 17, SD = 2.0).

The relationship between aircraft count and controller workload was examined across the three equipage levels. Repeated measures Two-Way Analysis of Variance (ANOVA) was used to analyze the Measurement Type (AC vs. workload) and Data Comm equipage (10%, 50%, and 90%). As expected, the results showed a significant difference between the average aircraft count (M = 17.2) and workload (M = 3.3; $F(1,7) = 194$, $p < 0.001$) and no difference across the Data Comm equipage conditions (M = 10, 10, 11 for 10%, 50%, and 90% respectively; $F(2,6) = 2.5$, $p > 0.1$).

More importantly, there was a significant interaction between the Measurement Type and the Data Comm equipage ($F(2,6) = 50.2$, $p < 0.001$), suggesting that the workload was significantly reduced because of an increase in Data Comm equipage even when the mean aircraft count increased (See Figure 4). The relationship between aircraft count and workload in this study illustrates why simple aircraft count may no longer be a valid metric as a workload predictor in a mixed equipage environment.

In this study, TBX modified the aircraft count to differentially weigh the Data Comm equipped aircraft to properly account for the workload associated with the equipage level. The TBX values showed a similar trend as the average workload ratings. The TBX values were the highest for the 10% Data Comm equipage condition (M = 22, SD = 2.0), lowest for the 90% condition (M = 20, SD = 1.5), and the 50% condition somewhere in between (M = 21, SD = 1.4). A repeated measures Two-Way ANOVA was used again but this time to analyze the relationship between TBX and workload. The analysis of Measurement Type (TBX vs. workload) and Data Comm equipage (10%, 50%, and 90%) showed a significant difference between the TBX values (M = 21.1) and workload (M = 3.3; $F(1,7) = 816$, $p < 0.001$), as expected.

More importantly, there was a main effect of Measurement Type (M = 13, 12, 11 for 10%, 50%, and 90% respectively; $F(2,6) = 17.4$, $p < 0.005$) but no significant interaction between the Measurement Type and the Data Comm equipage ($F(2,6) = .50$, $p > 0.6$), suggesting that TBX values mirrored the workload better than the simple aircraft count.

The relationship between workload, aircraft count, and TBX values was further examined using correlation analysis. As shown in Table 1, the correlation results were consistent with our hypothesis. There was a significant positive relationship between workload and TBX, $r(6) = 0.967$, $p = 0.002$, indicating that the TBX can be used to accurately reflect the controller workload. In contrast, the aircraft count had a significant negative relationship with workload, $r(6) = -0.913$, $p = 0.011$, and TBX, $r(6) = -0.869$, $p = 0.025$.

Table 1. Correlation between controller workload, aircraft count, and TBX

Pearson Correlation	Workoad	Aircraft Count	TBX
Workoad	1		
Aircraft Count	-0.913*	1	
TBX	0.967**	-0.869*	1
** Correlation is significant at the 0.01 level (2-tailed).			
* Correlation is significant at the 0.05 level (2-tailed).			

Although the TBX values reflected the workload fairly well and allowed the supervisors and TMCs to manage controller workload effectively, the TBX values in Figure 4 suggest that further tuning of the parameters is warranted. The workload rating differences between the conditions were substantial, ranging from 2.6 (easy to moderate) for the 90% condition and 4 (moderate to hard) for the 10% condition. In contrast, the TBX values did not show as noticeable a difference across the conditions. Greater weighting differences between the equipped and unequipped aircraft would further enhance TBX calculations in future simulations.

CONCLUSION

Aircraft count is used in current air traffic operations to predict controller workload in a traffic management context. However, aircraft count negatively correlated with controller workload in our simulated mixed equipage NextGen environment. This demonstrates that NextGen concepts require consideration of automation and new technologies in estimating controller workload.

Therefore TBX was introduced to determine relevant complexity factors and attribute proper workload levels to each factor. TBX uses just a few, easily understandable complexity factors that air traffic personnel readily cite as being major contributors to traffic complexity and controller workload (e.g., Hilburn and Flynn, 2005). Emerging NextGen aspects (e.g., Data Comm equipage mix) were easy to initialize and fine tune later within the TBX framework.

Overall, TBX calculations demonstrated a method of presenting traffic complexity and estimated controller workload for NextGen environments that are uniquely able to account for and reflect a diverse set of enabling technologies and the changing role of air traffic controllers.

REFERENCES

Bloem, M., Brinton, C., Hinkey, J., Leiden, K. and Sheth, K. (2009), A Robust Approach for Predicting Dynamic Density, *9th AIAA Aircraft Technology, Integration, and Operations (ATIO) Conference*, Hilton Head, SC.

Chatterji, G. B. and Sridhar, B. (2001), Measures for Air Traffic Controller Workload Prediction, *1st AIAA Aircraft Technology, Integration, and Operations (ATIO) Conference*, Los Angeles, CA.

Masalonis, A. J., Callaham, M. B., and Wanke, C. R. (2003), Dynamic Density and Complexity Metrics for Realtime Traffic Flow Management, *5th USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary.

Hilburn, B. and Flynn, G. (2005), Modeling Air Traffic Complexity: Eurocontrol's COCA Approach, *24th Digital Avionics Systems Conference*, Crystal City, Washington D. C.

Histon, J. M. and Hansman, R. J. (2008), Mitigating Complexity in Air Traffic Control: The Role of Structure-based Abstractions, Technical Report, ICAT-2008-05, Massachusetts Institute of Technology, Cambridge, MA.

Joint Planning and Development Office, (2007), Concept of Operations for the Next Generation Air Transportation System, version 2.0.

Kopardekar, P. and Magyarits, S. (2003), Measurement and Prediction of Dynamic Density, *5th USA/Europe Air Traffic Management R&D Seminar*, Budapest, Hungary.

Prevot, T. and Lee, P. U. (2011), Trajectory-Based Complexity (TBX): a Modified Aircraft Count to Predict Sector Complexity during Trajectory-Based Operations, *30th Digital Avionics Systems Conference*, Seattle, WA.

Smith, N., Prevot, T., Brasil, C., Homola, J., Kessell, A., Lee, H., Lee, P., Mainini, M., and Mercer, J. (2011), A Human-in-the-Loop Evaluation of Flow-Based Trajectory Management in Mixed Equipage Airspace, *Ninth USA/Europe Air Traffic Management Research and Development Seminar, ATM2011* (Paper 132), EUROCONTROL/FAA. Berlin, Germany, June 2011